

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-132449

(43)Date of publication of application : 12.05.2000

(51)Int.Cl.

G06F 12/00

G06F 13/00

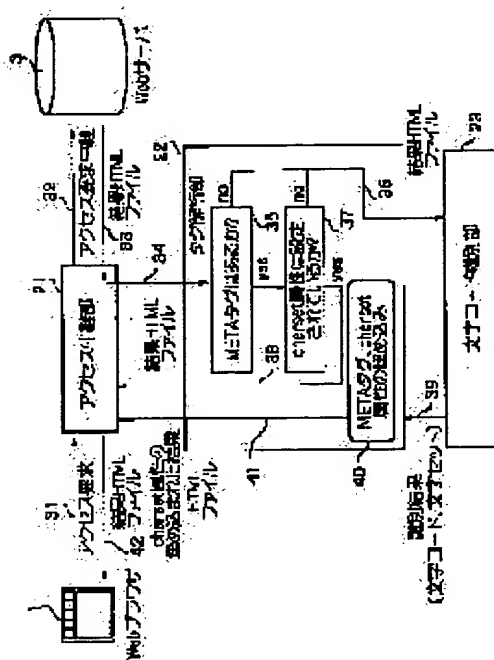
(21)Application number : 10-305472

(71)Applicant : NIPPON TELEGR &
TELEPH CORP <NTT>

(22)Date of filing : 27.10.1998

(72)Inventor : WATANABE KAZUNARI

(54) PROXY ACCESS METHOD, DEVICE THEREFOR AND RECORD MEDIUM RECORDED WITH PROXY ACCESS PROGRAM



(57)Abstract:

PROBLEM TO BE SOLVED: To enable a user to obtain a display free of illegal character without changing settings of a Web browser in a system comprising the Web browser, the proxy access device, and a Web server.

SOLUTION: A result HTML file from the Web server 3 after being received by an access repeating part 21 is given to a tag analysis part 22. The tag analysis part 22 received the result HTML file and checks whether or not there is a META tag. When no META tag is present, the HTML file is given to a character code discrimination part 23. When a META tag is found, the setting of charset (character set) properties in the tag is checked and when no setting is made, the result HTML file is given to the character code discrimination part 23. The character code discrimination part 23 checks the given HTML to discriminates its character code

and character set and gives the discrimination results to the tag analysis part 22. The tag analysis part 22 embeds the META tag and charset properties in the result HTML file and returns it to the Web browser through the access repeating part 21.

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-132449

(P2000-132449A)

(43)公開日 平成12年5月12日(2000.5.12)

(51)Int.Cl.⁷

識別記号

F I

テーマコード* (参考)

G 0 6 F 12/00
13/00

5 4 6
3 5 4

G 0 6 F 12/00
13/00

5 4 6 K 5 B 0 8 2
3 5 4 D 5 B 0 8 9

審査請求 未請求 請求項の数6 O L (全 6 頁)

(21)出願番号

特願平10-305472

(22)出願日

平成10年10月27日(1998.10.27)

(71)出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72)発明者 渡辺 一成

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74)代理人 100100893

弁理士 渡辺 勝 (外1名)

Fターム(参考) 5B082 EA07 GC03

5B089 GA11 GB03 GB04 JA22 JA24

JA32 JB02 KA09 KB04 KC11

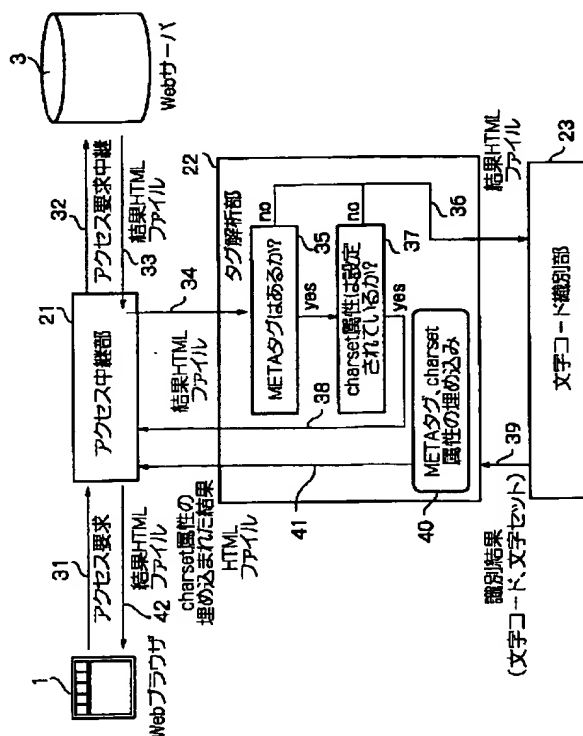
KC21 KC59 KF05 KH02 LB14

(54)【発明の名称】 代理アクセス方法、装置、および代理アクセスプログラムを記録した記録媒体

(57)【要約】

【課題】 Webブラウザ、代理アクセス装置、Webサーバからなるシステムにおいて、利用者がWebブラウザの設定を変更することなく文字化けのない表示を可能にする。

【解決手段】 Webサーバ3からの結果HTMLファイルはアクセス中継部21で受け付けられた後、タグ解析部22に渡される。タグ解析部22は結果HTMLファイルを受け取り、METAタグの有無のチェックする。METAタグが無ければ文字コード識別部23にHTMLファイルを渡す。また、METAタグがあれば、タグ中のcharset 属性の設定を調べ、設定されていなければ、結果をHTMLファイルを文字コード識別部23に渡す。文字コード識別部23は渡されたHTMLを検査し、その文字コード、文字セットを識別し、識別結果をタグ解析部22に渡す。タグ解析部22は識別結果をもとに結果HTMLファイルにMETAタグ、charset 属性を埋め込み、アクセス中継部21を介してWebブラウザ1に返す。



【特許請求の範囲】

【請求項1】 WebブラウザからWebサーバへのアクセス要求を受け付け、同要求を前記Webサーバに改めて発行し、その後Webサーバから返却されたファイルを一旦受信し、前記Webブラウザに返却する代理アクセス方法において、

文字コード、文字セットの情報が受信したファイルに明示的に記述されているか解析し、記述されていない場合には、前記文字コード、文字セットを識別し、識別結果を前記ファイル中に明示的に埋め込んだ後、該ファイルを前記Webブラウザへ返却することを特徴とする代理アクセス方法。

【請求項2】 前記ファイルがHTMLファイルであり、前記文字コード、文字セットがMETAタグ、charset属性である、請求項2記載の方法。

【請求項3】 WebブラウザからWebサーバへのアクセス要求を中継し、また前記Webサーバから返却された結果ファイルを受信し、結果ファイルを前記Webブラウザへ返却するアクセス中継部と、

前記受信された結果ファイルに文字コード、文字セットの情報が明示的に記述されているかどうか解析し、また識別結果を前記ファイル中に埋め込んだ後前記結果ファイルとして前記アクセス中継部に出力するタグ解析部と、

前記タグ解析部による解析の結果、文字コード、文字セットの情報がファイルに明示的に記述されていないと判断された場合、文字コード、文字情報の識別を行い、前記識別結果を前記タグ解析部に出力する文字コード識別部を有する代理アクセス装置。

【請求項4】 前記ファイルがHTMLファイルであり、前記文字コード、文字セットがMETAタグ、Charset属性である、請求項3記載の装置。

【請求項5】 WebブラウザからWebサーバへのアクセス要求を中継し、また前記Webサーバから返却された結果ファイルを受信し、結果ファイルを前記Webブラウザへ返却するアクセス中継処理と、

前記受信されたファイルに文字コード、文字セットの情報が明示的に記述されているかどうか解析し、識別結果を前記ファイル中に埋め込んだ後前記結果ファイルとして前記アクセス中継処理に出力するタグ解析処理と、前記タグ解析処理による解析の結果、文字コード、文字情報のセットがファイルに明示的に記述されていないと判断された場合、文字コード、文字情報の識別を行い、前記識別結果を前記タグ解析処理に出力する文字コード識別処理をコンピュータに実行させるための代理アクセスプログラムを記録した記録媒体。

【請求項6】 前記ファイルがHTMLファイルであり、前記文字コード、前記文字セットがMETAタグ、charset属性である、請求項5記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、WebブラウザからWebサーバへのアクセス要求を受け付け、同要求をWebサーバに改めて発行し、その後Webサーバから返却されたファイルをWebブラウザに返却する代理アクセス方法および装置に関する。

【0002】

【従来の技術】Webブラウザでは利用者が予め設定した文字コード、文字セット（文字の集合）（またはデフォルトで定められているそれ）にしたがい受信したファイルの表示を行う。設定と実際に受信したファイルの文字コード、文字セットが異なる場合には文字化けとなるが、利用者がそのつど設定を変更することで対処する。

【0003】また、従来の代理アクセス装置はWebブラウザからWebサーバへのアクセスの中継を行う際に、受信したファイルの文字コード変換を行った後にWebブラウザに返却する。この技術では元のファイルの文字コード、文字セットの種類に係わらず、ある決まった文字コードに変換してWebブラウザに返却する。Webブラウザ側で予め変換後の文字コードを設定しておくことで文字化けを回避する。

【0004】

【発明が解決しようとする課題】Webブラウザでは上記従来の技術で述べたように設定と実際のファイルの文字コード、文字セットが異なる場合、利用者がその都度設定を変更する必要がある。多言語環境でのWebブラウザでは利用者がその設定を変更する状況が頻繁に発生し、その設定負荷が大きな問題となる。そこで設定負荷を極力減らすような機構が必要である。

【0005】また、従来の代理アクセス装置はWebサーバから受信したファイルを決まった文字コードに変換するが、この方法はファイルの文字コード、文字セットを意識せず文字コード変換を行うため閲覧対象となるファイルが多言語である場合うまく動作しない。例えば、代理アクセス装置が全てのファイルをSJISに変換すると仮定する。閲覧対象の文字コードが日本語のJISファイル、ASCIIファイルである場合はそれらをSJISに変換することは可能であるのでWebブラウザは変換後のファイルを正常に表示できる。しかしながら閲覧対象に韓国語や中国語のEUCファイルが含まれているとそれらをSJISに変換することはできないので変換後のファイルを受け取ったWebブラウザでは文字化けする。さらに問題となるのは韓国語EUC（Extended Unix Code）ファイルであっても無理やりSJISに変換してしまうので、変換後のファイルの文字コードは規定されていないものになってしまう。したがって、Webブラウザ側の設定をいくら変更してもそのファイルは永久に閲覧することはできない。

【0006】Webサーバから返却される全てのファイルに文字コード、文字セットが明示的に指定されていれ

ば上記のような問題は発生せず、Webブラウザはその指定にしたがい表示するだけで文字化けは回避できる。しかしながら、実際のファイルでそのような指定はほとんどされておらず、Webブラウザ側でその情報が得られないために上記のような問題が生じることになる。例えば、HTML (Hyper Text Markup Language) ではMETAタグが文字コード、文字セットを指定するためのタグである。METAタグはその属性として、charset (character set) 属性を持つことができ、その属性値で文字コードと文字セットを指定できる。これらはそのHTMLファイルがどのような文字コード、文字セットで記述してあるかを明示的に示すためのものである。WebブラウザはMETAタグとcharset 属性があれば利用者の設定に頼らずその文字コード、文字セットにしたがった表示を行う。しかし、現状のWebサーバ中のHTMLファイルにおいてこのMETAタグとcharset 属性を指定してあるものはごく少数であり、Webブラウザはその情報を有効に利用することができない。Webブラウザはその情報を補うために利用者の設定する文字コード、文字セット情報を利用している。

【0007】本発明の目的は、利用者がWebブラウザの設定を変更することなく文字化けのない表示が可能な代理アクセス方法、装置、および代理アクセスプログラムを記録した記録媒体を提供することにある。

【0008】

【課題を解決するための手段】Webブラウザ、代理アクセス装置、Webサーバからなるシステムにおいて、Webブラウザからのアクセス要求は一旦代理アクセス装置で受け付けられ、改めて代理アクセス装置からWebサーバへ発行される。Webサーバから返却されたファイルは代理アクセス装置で一旦受信され、当該ファイル中の文字コード、文字セット指定のための情報の有無のチェックを行う。設定されていなければ当該ファイルの文字コード、文字セットの識別を行い、識別情報を付加した後に、Webブラウザにファイルを返却する。Webブラウザでは指定情報にしたがった表示を行うため、多言語環境のブラウザにおいても、ブラウザの設定を変更することなく文字化けのない表示を行うことが可能である。また万が一、代理アクセス装置における識別に失敗し、ブラウザ側で文字化けしても従来の代理アクセス装置と異なりファイル自体のコード変換を行っていないのでブラウザ側の設定変更で表示が可能になる。

【0009】HTMLの場合はMETAタグのcharset 属性を利用し、その指定がない場合にはその情報をファイル中に埋め込んでWebブラウザに返却する。これによりWebブラウザに返却されるHTMLファイルには全てMETAタグ、charset属性が付与され、その情報に基づいた表示が可能となる。

【0010】

【発明の実施の形態】次に、本発明の実施の形態につい

て図面を参照して説明する。

【0011】図1を参照すると、本発明の一実施形態のシステムはWebブラウザ1と代理アクセス装置2とWebサーバ3から構成される。代理アクセス装置2はアクセス中継部21と解析部22と文字コード識別部23を含む。

【0012】図2は図1の実施形態の処理の手順を示す図である。本例はWebサーバ3からHTMLファイルが返却される場合である。利用者はWebブラウザ1で目的とするWebサーバ2のURL (Uniform Resource Locator) を格納することでアクセス要求を出す (ステップ31)。このとき、ブラウザのプロキシ指定を代理アクセス装置2に設定することで利用者は代理アクセス装置2経由のWebサーバ3の利用が可能になる。アクセス中継部21はブラウザ1からのアクセス要求を一旦受け付け、同じ要求を改めてWebサーバ3に対して発行する (ステップ32)、Webサーバ3からは結果HTMLファイルがアクセス中継部21で一旦受け付けられ、タグ解析部22に渡される (ステップ33, 34)。タグ解析部22はWebサーバ3から返却された結果HTMLファイルを受け取り、まずMETAタグの有無のチェックを行う (ステップ35)。METAタグが設定されていなければ当該HTMLファイルの文字コード、文字セットは明示的に指定されていないことになり、文字コード識別部23に渡す (ステップ36)。METAタグが設定されている場合、タグ中のCharset 属性の指定の有無をチェックする (ステップ37)。charset 属性が指定されている場合、明示的に当該HTMLファイルの文字コード、文字セットが指定されていることとなり、そのままアクセス中継部21に当該HTMLファイルを返却する (ステップ38)。charset 属性が指定されていない場合は当該HTMLファイルの文字コード、文字セットは明示的に指定されていないことになり文字コード識別部23による識別を行う (ステップ36)。

【0013】文字コード識別部23では渡されたHTMLファイルに対して検査を行い、その文字コード、文字セットを識別する。この処理には例えば特開平9-212334に記述されている方法が利用できる。識別結果の文字コード、文字セットの情報はタグ解析部22に返却される (ステップ39)。

【0014】タグ解析部22では結果HTMLファイル中にMETAタグ、charset 属性を追加した後にアクセス中継部21に返却する (ステップ40, 41)。図3(1)、図3(2)にHTMLファイル中に埋め込まれたMETAタグ、charset 属性の例を示す。

【0015】代理アクセス装置2で処理の終わった結果HTMLファイルはWebブラウザ1に返却される (ステップ42)。

【0016】最終的にWebブラウザ1に返されるHT

MLファイルは全てMETAタグとそのcharset 属性により文字コード、文字セットが明示的に指定されるため、Webブラウザ1は指定にしたがい結果HTMLファイルを表示することで文字化けすることない表示が可能となる。

【0017】図4を参照すると、本発明の一実施形態の代理アクセス装置は受信装置51と記憶装置52と送信装置53と記録媒体54とデータ処理装置55で構成されている。受信装置51はWebブラウザ1からのアクセス要求を受信し、またWebサーバ3から結果HTMLファイルを受信する。送信装置53はアクセス要求をWebサーバ3へ送信し、また結果HTMLファイルをWebブラウザ1へ送信する。記憶装置52はハードディスクである。記録媒体54は図2に示した代理アクセス装置2の処理が代理アクセスプログラムとして記録された、フロッピー・ディスク、CD-ROM、光磁気ディスク等の記録媒体である。データ処理装置55は記録媒体54から代理アクセスプログラムを記憶装置52に読み込んで、これを実行するCPUである。

【0018】

【発明の効果】以上述べたように、本発明によれば、利用者がWebブラウザの設定を変更することなく文字化けのない表示が可能であり、設定のための負荷を大幅に

低減することができる。

【図面の簡単な説明】

【図1】本発明の一実施形態のシステム構成図である。

【図2】図1の実施形態における処理の流れを示す図である。

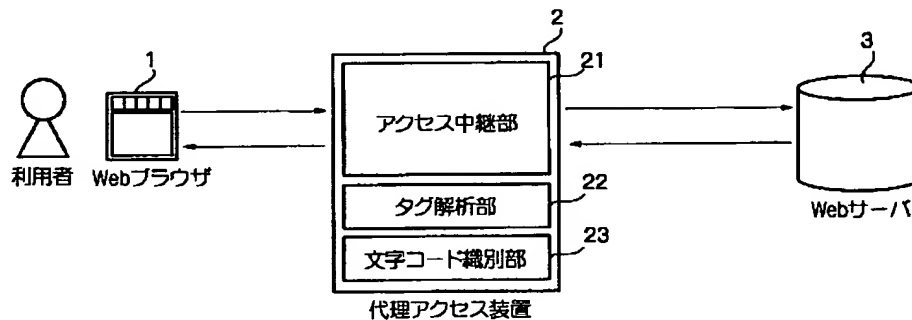
【図3】HTMLファイル中に埋め込まれたMETAタグとCharset 属性の例を示す図である。

【図4】代理アクセス装置の他の実施形態を示す構成図である。

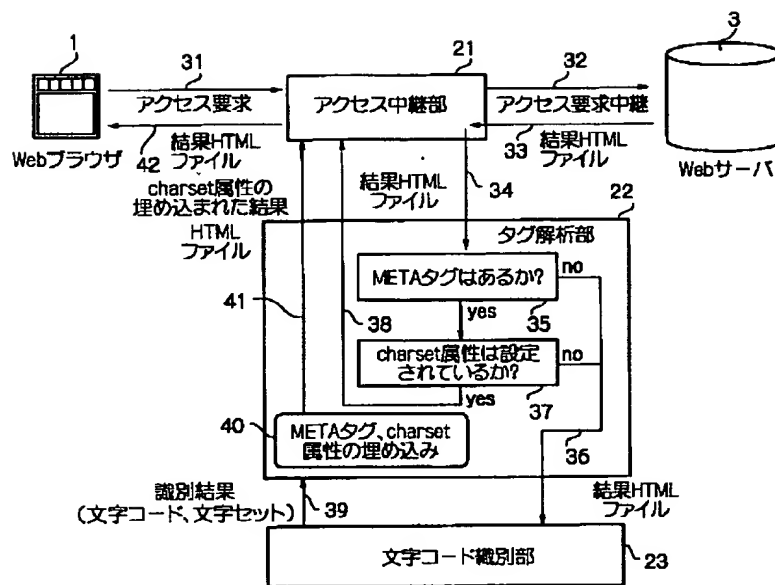
【符号の説明】

- | | |
|-------|----------|
| 1 | Webブラウザ |
| 2 | 代理アクセス装置 |
| 3 | Webサーバ |
| 21 | アクセス中継部 |
| 22 | タグ解析部 |
| 23 | 文字コード識別部 |
| 31～42 | ステップ |
| 51 | 受信装置 |
| 52 | 記憶装置 |
| 53 | 送信装置 |
| 54 | 記録媒体 |
| 55 | データ処理装置 |

【図1】



【図2】



【図3】

(1)

```
<HTML>
<HEAD>
<TITLE>TEST1</TITLE>
<METAHTTP-EQUIV="Content-Type"CONTENT="text/html;charset=x-euc-jp">
</HEAD>
```

(2)

```
<HTML>
<HEAD>
<TITLE>TEST1</TITLE>
<META charset=x-euc jp>
<HTML>
</HEAD>
```

【図4】

